Bayes decision rule

什么是贝叶斯公式

已知事件A,B发生的概率分别为 P(A), P(B), 那么两者的相互的条件概率(conditional probability)分别可以表示为: P(A|B) , P(B|A).

P(A|B) = $\frac{P(A\&B)}{P(B)}$ 得出 P(A & B) = P(A|B) * P(B)

P(B|A) = $\frac{P(A\&B)}{P(A)}$ 得出 P(A & B) = P(B|A) * P(A)

所以可以推出 => P(A|B) * P(B) = P(B|A) * P(A) 稍加变形就是**贝叶斯公式** :

P(A|B) = $\frac{P(B|A)*P(A)}{P(B)}$

贝叶斯公式的小例子

假设以硬币举例

- 假设已知硬币正面, 反面朝上的概率分别为 P(正) P(反) 叫做 **先验概率((Priori Probabilities)**
- 在已知先验概率的情况下, 求在硬币为某一属性 *X* 下, 正面的概率为 : P(正 | 硬币 = X) 叫做后验概率(**posterior probability**)

例如, 我们现在抛硬币, 且知道硬币有三种类型(数据属性) c_1 = 大 , c_2 = 中 , c_3 = 小 , 得到的结果为正, 反(输出类型), 其中30次反面, 70次正面. 在反面时, 小硬币出现6次. 正面时, 小硬币出现7次

所以我们的任务就是求的后验概率 **当观察到硬币为小硬币, 其结果分贝为正反的概率**, 根据贝叶斯公式 :

P(正 | 硬币 = 小) = $\frac{P(\text{硬币}=\text{小}|\text{正面})*P(\text{正面})}{P(\text{硬币}=\text{小})}$

P(硬币 = 小) = $\frac{13}{100}$ = 0.13

P(硬币 = 小 | 正) = $\frac{7}{70}$ = 0.1

P(正) = $\frac{70}{100}$

所以可以求出概率 = 0.54 , 说明**当硬币是小硬币的情况下, 它抛出正面的概率为0.54**

贝叶斯方法在机器学习分类中的应用

这种方法假设我们已经完全知道了每个分类的 **先验概率(Priori Probabilities)** , 即我们知道如果结果有 *C* 类, 那么P(c_1) , ... , P(c_c) 我们都知道.

这种方法应用到机器学的分类中本质就是需要: 根据先验概率(建模时获得), 在观察到新的数据的属性X时候, 求的它的后验概率的过程

bayes 算出所有类的后验概率

即对于所有的 {c_1, ..., c_w} $\in$ Class :

然后根据所有的后验概率找到最大的那一个 Max P(c_w | X), 记为该数据属性的分类 c_w

P(c_i | X) = $\frac{P(X|c_i)*P(c_i)}{P(X)}$

其中 P(X) 的概率都一样 : P(X) = $\sum_{j=1}^{C} P(X|c_i) * P(c_i)$

所以有了结论, **我们将X归于类c_i当且仅当** (Bayes' rule minimum error):

$P(x|c_i) * P(c_i) > p(x|c_k) * P(c_k)$ **k = 1, ..., w k$\neq$ i**

Bayes 方法优缺点

优点

- 贝叶斯决策理论是最优的, 因为只要观测到数据属性为 *X* , 然后选择后验概率最大的结果, 就可以**最小化预测错误的概率(Bayes decision rule for minimum risk)可以证明.** 这个结论对所有的观测值 $X \in$ All_features 都成立 , 从而可以保证预测错误的概率最小, 从而达到最优.
- **可以调节先验概率和观测现象之间的平衡**, 即有了Bayes方法可以使得预测结果不仅仅依赖于先验概率, 还一部分取决于观测的现象(数据属性 *X*)

缺点

- 我们通常是无法获得先验概率(prior class), 条件概率(conditional densities)等计算要件. 要获得只有从数据中进行估算(比如上面抛硬币的例子), 所有肯定和真实的概率分布有误差.
- 往往在实际中, 要观察的特征不知一个(不仅仅为大中小), 可能会面临成百上千个特征属性. 所以在计算的时候, 会遇到**维度灾难(the curse of dimensionality)**, 时候计算的数值很不稳定. 所以才会发展出**朴素贝叶斯方法的平滑(smoothing)**去解决这个问题